A Hybrid Model For Short-Term Traffic Volume Prediction In Massive Transportation Systems

Zulong Diao, Dafang Zhang, Xin Wang, *Member, IEEE*, Kun Xie, *Member, IEEE*, Shaoyao He, Xin Lu, and Yanbiao Li

***Abstract*— The prediction of short-term volatile traffic becomes increasingly critical for efficient traffic engineering in intelligent transportation systems. Accurate forecast results can assist in traffic management and pedestrian route selection, which will help alleviate the huge congestion problem in the system. This paper presents a novel hybrid DTMGP model to accurately forecast the volume of passenger flows multi-step ahead with the comprehensive consideration of factors from temporal, origin-destination spatial, and frequency and self-similarity perspectives. We first apply discrete wavelet transform to decompose the traffic volume series into an appropriation component and several detailed components. Then we propose a more efficient tracking model to forecast the appropriation component and a novel Gaussian process model to forecast the detailed components. The forecasting performance is evaluated with real-time passenger flow data in Chongqing, China. Simulation results demonstrate that our hybrid model can achieve on average 20%–50% accuracy improvement, especially during rush hours.**

***Index Terms*— Passenger flow prediction, wavelet decomposition, Gaussian process (GP).**

I. Introduction

THE accelerated urbanization process and urban population explosion bring great pressure to the urban traffic management. In order to withstand the high traffic pressure especially during the morning rush hours or holidays, many cities have been planning and building rail transit systems. In addition, smart cards are promoted to speed up the traffic flow in railway stations and facilitate traffic management.

Manuscript received February 5, 2017; revised July 12, 2017, November 22, 2017, and March 13, 2018; accepted May 6, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61472130 and Grant 61572184, in part by the Planned Science and Technology Project of Hunan Province, and in part by the Hunan Provincial Natural Science Foundation of China under Grant 2017JJ1010. The Associate Editor for this paper was V. Punzo. *(Corresponding authors: Dafang Zhang; Kun Xie; Shaoyao He.)*

Z. Diao, D. Zhang, K. Xie, and Y. Li are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: diaozulong1988@sina.cn; dfzhang@hnu.edu.cn; xiekun@hnu.edu.cn; lybmath_cs@hnu.edu.cn).

X. Wang is with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794 USA (e-mail: x.wang@stonybrook.edu).

S. He is with the College of Architecture, Hunan University, Changsha 410082, China (e-mail: syhe829@163.com).

X. Lu is with the School of Business, Central South University, Changsha 410083, China, and also with the Department of Public Health Sciences, Karolinska Institutet, 17177 Stockholm, Sweden (e-mail: xin.lu@flowminder.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2018.2841800

A well-designed transportation system is a key element in the economic welfare of major cities [1]. For the efficient operation of urban transit system, planners must have a quantitative understanding of traffic patterns, and be able to predict and prevent any disruptions, either planned or unplanned [2]. More specifically, to enable real-time traffic management of subway operations and support functions such as regulating station passengers, it is important to accurately forecast the volume of station passenger flows. Passengers enter the light rail system from the origin station and leave the system from the destination station. Given an origin-destination (OD) station pair in a light rail system, the key questions that need to be answered are: (1) How to predict the number of passengers that exit from the destination station at time t? (2) How to improve the average prediction accuracy per day? (3) How to improve the prediction accuracy at rush hours so as to find the crowded moment that passenger flows reach the peak and become much greater than the historical average?

Traffic volume forecasting is fundamental to the performance of many components in intelligent transportation systems. Great efforts have been devoted to improve the forecasting accuracy and the transportation efficiency. Designed based on purely spatial or purely temporal information, the forecasting performance of most known approaches [3]–[24] is often low. The traffic patterns in the massive transportation systems are affected by various factors. Several recent studies [25]–[31] try to improve the traffic forecasting accuracy by decomposing the traffic data into different frequency components to understand the flow evolution from the frequency perspective. Hybrid methods are also proposed to identify the daily traffic patterns and some other flow patterns to improve the forecasting accuracy [32]–[34].

Although the approaches based on the frequency analysis present good performance on capturing the overall traffic trend, most of them fail to track the local fluctuation of passenger flows. High frequency sub-signals extracted by the frequency domain analysis are still fitted by traditional non-linear models, which can't extract enough information to make accurate prediction. In addition, very limited studies are made on the relationship between an OD flow between a pair of stations and the overall passenger flow leaving from the same origin. Our preliminary studies indicate that the OD flow often has a high volatility when the corresponding passenger flow exhibits intense oscillations or steeply changes.

In this work, we propose a novel hybrid model to more accurately forecast the volume of passenger flows multi-step ahead with the comprehensive consideration of factors from temporal, OD spatial, frequency and historical probabilistic distribution perspectives. Our contributions are summarized as follows:

- To extract flow features from both the temporal and frequency domains, we apply discrete wavelet transform DWT) to decompose the traffic volume series into a set of sub-signals.
- To deal with the chaos and uncertainty of traffic flow, we apply the state space reconstruction to preprocess the detailed components extracted by DWT. Our analysis of real passenger flow data reveal that traffic time series belong to a type of chaotic signals, which are commonly handled with the state space reconstruction method.
- To capture temporal and spatial OD features, we propose a more efficient tracking model to timely adapt the predicted value of exit passenger flow according to the entrance passenger flow.
- To better follow the dynamic patterns of flows, we propose a novel Gaussian Process (GP) model to forecast the detailed components from the probabilistic perspectives. We exploit the simultaneous predictions, with each based on the state space reconstructed with a different time delay between two adjacent components of the state vector.
- To evaluate the performance of our proposed hybrid model, we have performed extensive simulations based on real passenger flow data. Simulation results demonstrate that the proposed hybrid model could significantly improve the prediction accuracy, especially in rush hours.

The rest of this paper is organized as follows. Section II summarizes the related work on the recent popular techniques for the prediction of traffic volume. The introduction of the background and the analysis of data are given in Section III and Section IV, respectively. Section V presents the technical details of our hybrid prediction model. We evaluate the performance of our proposed model through simulations driven by the real-world data sets in Section VI, and conclude our work in Section VII.

II. Related Work

In this section, we review the related work on traffic volume forecasting, and identify the differences of our work from the existing work.

Existing traffic forecasting techniques can be mainly divided into two categories, parametric and non-parametric [3], [4]. The common parametric techniques include the classical Autoregressive Moving Average (ARIMA) [5]–[8] and Kalman filtering model [9], [10]. Because of the fast and easy operation, they are widely used in transportation systems in the early stages. Due to the stochastic and nonlinear nature of traffic flow, researchers have also paid much attention to nonparametric methods in the traffic flow forecasting field. Wu *et al.* [14] utilize support vector regression (SVR) with a radial basis function kernel to predict the travel time in a transportation network. In [15] and [16], an online version of SVR is employed to solve the same problem. Oswald *et al.* [17] use the nearest neighbor approach to forecast the traffic flow. Yu *et al.* [18] apply Gaussian mixture model and expectation-maximization to estimate the density functions of transition probability and predict traffic flow. Lv *et al.* [19] propose the deep-learning-based traffic flow prediction, which considers the spatial and temporal correlations inherently. Methods proposed in [20]–[24] attempt to capture the process dynamics and improve forecasting accuracy based on GP model. These methods can generally capture the long-term trend of the traffic flow, but fail to capture the traffic short-term fluctuations while the overload is one major factor that impacts the performance of transportation systems.

In recent years, a number of approaches are proposed to better capture the traffic features by decomposing the traffic data into different components. Hu and Wang [25] and Agarwal *et al.* [26] apply discrete wavelet transform (DWT) to make short-term wind speed prediction and detect traffic incidents. Wei and Chen [27] develop a hybrid EMD-BPN forecasting approach that combined empirical mode decomposition (EMD) with back-propagation neural networks (BPN) to predict short-term passenger flow in metro systems. Wang *et al.* [28] propose a least squares support vector regression ensemble learning model based on the seasonal decomposition for the forecast of the Chinese hydropower consumption. Xing *et al.* [29] introduce the Robust Principal Component Analysis for passenger flow decomposition. This method can more accurately and robustly capture the underlying temporal and spatial characteristics of passenger flow in the presence of all kinds of fluctuations. Zhang *et al.* [30], [31] present a hybrid model for multi-step ahead passenger flow forecasting based on spectral analysis technique, ARIMA model and generalized autoregressive conditional heteroskedastic model. Decomposing traffic data into different components can help more accurately track the temporal evolution of the traffic flow at different time resolutions and improve traffic forecasting accuracy. However, the performance of existing approaches is often low and not stable. Most existing studies use the same method to handle different components without considering their difference, and simply fit the high frequency components with traditional non-linear models that cannot follow well the sophisticated behaviors of transportation systems.

Some recent studies attempt to apply hybrid methods that can take advantage of different models to identify both the common daily traffic trend and some variations in flow patterns to improve the forecasting accuracy. Wang and Shi [32] propose a traffic speed forecasting hybrid model based on support vector machine (SVM), DWT and phase space reconstruction. In [33], a hybrid modeling approach which combines artificial neural networks with a simple statistical is studied to make prediction for traffic flow. Silva *et al.* [34] propose a new approach to analyze massive transportation systems that used data obtained from smart cards in the London transport system to predict future passenger flow. These methods can achieve higher overall forecasting accuracy, but they often fail to identify the time of congestion or fail to accurately predict the congestion degree in rush hours. This is because these models still cannot extract enough information from traffic flow data for forecasting.

In this paper, we design a novel hybrid method with the comprehensive consideration of several factors, including temporal and frequency signal features, spatial relationship among OD pairs, and probabilistic evolution of passenger flows. Therefore, we can extract more valuable information from flow data to improve the forecasting accuracy. Different from previous studies, we deal with the approximation component and detailed components generated by DWT in different ways. To capture temporal and OD spatial features, we apply a more effective tracking model to predict the approximation component. We further propose a mixed time delay Gaussian process (MTGPR) model to predict detailed components with the intention of achieving the least predicted variance. Gaussian Process models are nonparametric kernel-based probabilistic models

III. Preliminaries

Before presenting our detailed design for more accurate forecast of passenger flows, we provide some background knowledge on related techniques.

A. State space reconstruction

The number of customers in the transportation system often varies over time due to the impact of factors such as climate, weather, holidays, peak hours and events. These factors, either expected or uncertain, could lead to complicated changes of the customer flows. These changes are hard to be tracked by a conventional linear system.

For the analysis of non-linear time series, Froehling *et. al* [35] introduce the concept of phase space reconstruction and the chaos theory. Based on the theory, all the dynamical information needed for determining a system state is included in the time series of any a system variable. A state trajectory constructed from a single-variable time series $\{y_i, i = 1, 2, \cdots, N\}$ of the length N maintains the foremost characteristics of the state trajectory in the original space. The reconstructed phase space can be represented as

$$x_i = (y_i, y_{i+\tau}, \cdots, y_{i+(d-1)\tau}), \quad i = 1, 2, \cdots, N - (d-1)\tau \tag{1}$$

where x_i is a state in the reconstructed phase space, d is the embedding dimension, and τ is the delay between adjacent elements of a state in the reconstructed phase space.

According to the Takens Theorem, if $d \geq 2D + 1$, where D is the dimension of the dynamic system, the reconstructed dynamic system and the original one are topologically equivalent. The system state of the next moment x_{i+1} can be predicted from the current state of the system. That is, acquiring the predication value for the next instant of the time series provides a basis for predicting chaos time series.

Determining an acceptable minimum embedding dimension d and the time delay τ between measurements for the state space reconstruction is crucial [36], [37]. Given an appropriate value of d, it is difficult to ensure that a chosen value of the delay τ is optimal all the time. As reported from past studies, if τ is too small, the system's dynamical characteristics will not be revealed; if it's too big, a simple trajectory will be made complicated, which can reduce the number of effective data points. We will discuss in Section VI how to make use of multiple delay values to increase the prediction accuracy.

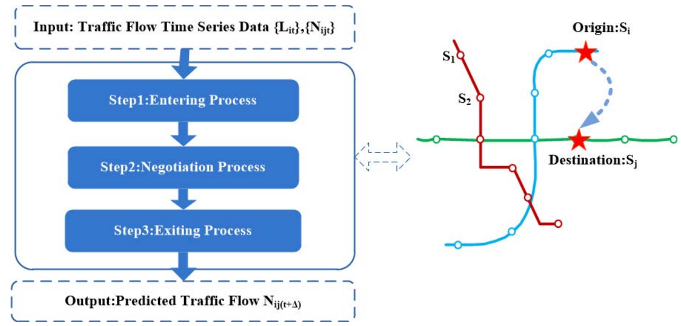


Fig. 1. Schematic of passenger flow prediction with a tracking model.

B. Gaussian process regression (GPR)

Gaussian process regression is an applicable and practical probabilistic approach based on statistical learning theory and Bayesian theory, which is widely used for state and performance prediction in a variety of fields. A Gaussian process model seeks to establish a mapping f of the form

$$y = f(\boldsymbol{x}) \tag{2}$$

between the predictor (output) $y \in R$ of a complex dynamical system and the input vector $\boldsymbol{x} \in R^d$. It assumes that any finite set of function values have a joint Gaussian distribution. Let $\boldsymbol{f}$ be the known function values of the training cases, and let $\boldsymbol{f_*}$ be a set of function values corresponding to the test set inputs X_*. We write out the joint distribution as:

$$\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{f_*} \end{bmatrix} \sim N(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu_*} \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix}) \tag{3}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\mu_*}$ are the means for the training set and test set, K is the covariance for the training set, K_* the covariance for the training-test set, and K_{**} the covariance for the test set. Since we know the values for the training set $\boldsymbol{f}$, we are interested in the conditional distribution of $\boldsymbol{f_*}$ given $\boldsymbol{f}$, which is expressed as:

$$\boldsymbol{f_*}|\boldsymbol{f} \sim N(\boldsymbol{\mu_*} + K_*^T K^{-1}(\boldsymbol{f} - \boldsymbol{\mu}), K_{**} - K_*^T K^{-1} K_*) \tag{4}$$

C. Tracking Model

A tracking model can be established to follow the evolution of the traffic flow of an OD pair based on the flow features extracted from the traffic data. Silva *et al.* [34] developed the general model to forecast the traffic volume of the London transportation system, where the traffic flows are for fast train commuting service within the Greater London area. Tracking model was designed to predict three unknowns: (i) entering counts (entering process), (ii) the rate at which passengers

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

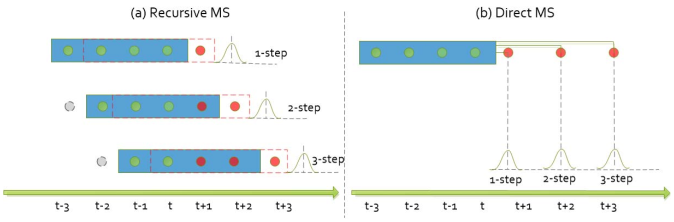


Fig. 2. Schematic of multi-step-ahead (MS) prediction with a GPR model.

remain inside the transportation system given these counts (negotiation process), and (iii) the rate at which passengers exit given the number of passengers inside the system and the length of their stay, according to the origin (exiting process). In Figure 1, L_{it} represents the number of passengers entering the station S_i at time t. N_{ijt} represents the number of passengers that have entered the system from the station S_i and exits (tapping-out) from the station S_j at time t.

D. Multi-Step Strategy

In many scenarios, there is a need to predict the system condition multiple steps ahead, and the Equation (5) shows the prediction for h steps. Recursive and Direct strategies [24], as shown in Figure 2, are two main approaches used for the multi-step ahead (MS) prediction. In the Recursive strategy, a single-step prediction model is trained by minimizing the square sum of the prediction errors to perform the first step prediction. Then the window of historical data moves ahead to include the newly predicted value and exclude the oldest sample. The process will continue until the step size of the multi-step prediction is reached. As the predicted values from past steps are always taken into the model as input to get the next prediction value, the prediction error is accumulated, thus largely affecting the prediction accuracy.

$$y_{t+h} = f_h(y_t, \cdots, y_{t-d+1}) + \varepsilon \tag{5}$$

Unlike the Recursive strategy, the window of historical data in the Direct strategy doesn't need to slide in every step of the process. However, to perform h-step ahead prediction, there is a need to train h models, each corresponding to an i-step ($i \in [1, h]$) prediction. This will incur a high computation overhead when h is large. Furthermore, because the prediction tasks for different ahead steps are run independently, the direct method may result in broken curves, especially when the system is nonlinear and dynamic.

IV. Empirical Study With Passenger Flow Data

A dynamic system is considered to be chaotic if the distance between two system evolution trajectories starting from very close initial position grows exponentially apart. Before presenting our model for prediction, we perform a set of experiments to investigate and identify the traffic patterns, especially the chaotic characteristics hidden in passenger flow series. From previous studies, we find that the chaotic characteristics often exist in the signals generated by the complex dynamic systems, such as the transportation system. As a result of the chaotic characteristics, signals can be computational unpredictable and sensitive to initial conditions. The dimension of the saturation correlation and the largest Lyapunov exponent are two fundamental measures of chaotic characteristics. If the chaotic characteristics really exist in traffic flow data, we need to carefully consider their impacts in our new model.

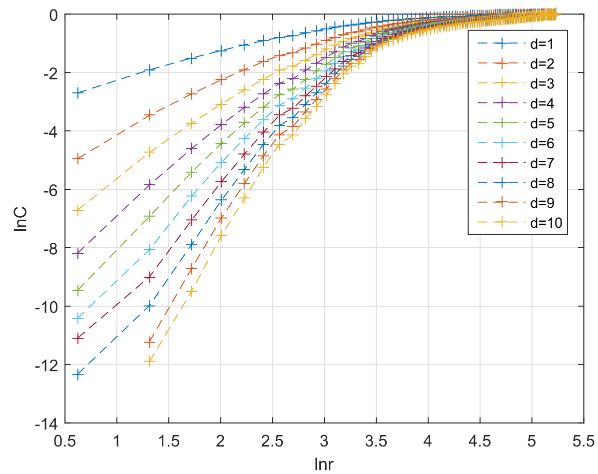


Fig. 3. $\ln C(r)/\ln(r)$ curve of passenger flow series.

We analyze the smart-card readings of light rail system in Chongqing, China. The data were collected over 150 days, from March 2014 to July 2014. Each reading consists of a card ID, a "tap-in" time stamp, a "tap-in" location code, a "tap-out" time stamp and a "tap-out" location code. A location code uniquely identifies one of 142 active stations of the system. To preserve the customer privacy, card IDs are anonymized and ignored in our analysis. The time resolution of the recorded time stamps is 1 min. Each day is composed of 1,200 mins, starting from 5:00 AM to 1:00 AM of the next calendar day. Weekdays are assumed to be exchangeable [34].

1) Calculation of the Correlation Dimension: Grassberger P [38] is a common method used for calculating the correlation dimension of the time series attractor. With the increase of the embedding dimension, the correlation dimension of random sequences increases but does not reach the saturation, while the correlation dimension of chaotic sequences approaches the saturation. Therefore, the chaotic sequence and the random sequence can be distinguished based on whether the correlation dimension has saturation phenomenon. The correlation dimension can be calculated as

$$D_2 = \lim_{d \to \propto, r \to 0} [\frac{\ln C_d(r)}{\ln(r)}]. \tag{6}$$

where $C(r)$ is the correlation function, r is the radius of the sphere in the reconstructed state space, d is the embedding dimension of the reconstructed state space. As shown in Figure 3, the curve $\ln C(r)$ versus $\ln(r)$ gradually becomes parallel (i.e., the slope of $\ln C(r)/\ln(r)$ becomes constant), with the increase of the embedded dimension d, which indicates that the correlation dimension gradually gets saturated and the passenger flow sequence has chaotic characteristics.

2) Calculating the Largest Lyapunov Exponents From Small Data Sets: The detection of the presence of chaos in a

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

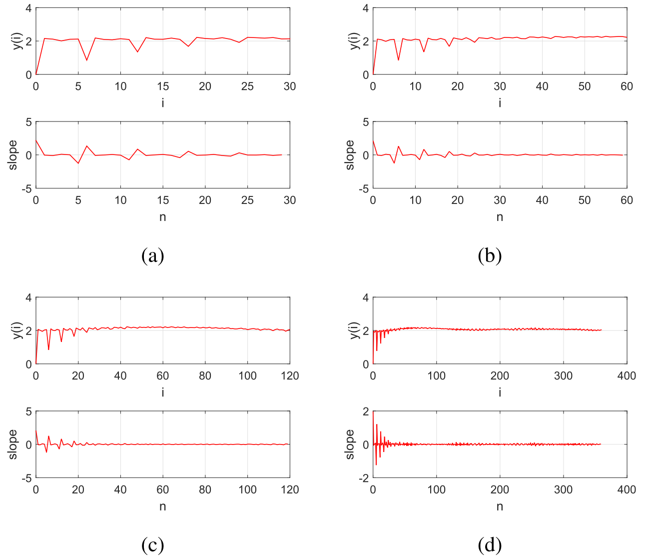


Fig. 4. The largest Lyapunov exponent of passenger flow series given different evolution step i: (a) i = 30, (b) i = 60, (c) i = 120, (d) i = 360.

dynamical system can be pursued by measuring the largest Lyapunov exponent. Lyapunov exponent quantifies the exponential divergence of initially close state-space trajectories and estimate the amount of chaos in a system. A practical method for calculating largest Lyapunov exponents is through algorithm for the small data sets [39]. We calculate the largest Lyapunov exponent as follows:

$$d_j(0) = \min_{X_{\hat{j}}} \left\| X_j - X_{\hat{j}} \right\|, \left| j - \hat{j} \right| > P$$
$$d_j(i) = \| X_{j+i} - X_{\hat{j}+i} \|, \quad i = 1, 2, \cdots, min(M - j, M - \hat{j}) \tag{7}$$

$$y(i) = \frac{1}{q\Delta t} \sum_{j=1}^{q} \ln d_j(i) \approx \frac{1}{q\Delta t} \sum_{j=1}^{q} (\ln d_j(0) + \lambda(i \cdot \Delta t)) \tag{8}$$

where i represents the discrete time-step, P represents the average time interval between two trajectories in the reconstructed phase space, X_j is an arbitrary point in the reconstructed phase space, Δt is the sample period, $d_j(0)$ is the shortest initial distance between two points in the reconstructed phase space, $d_j(i)$ is the distance between X_{j+i} and $X_{\hat{j}+i}$, q is the number of non-zero $d_j(i)$ values and $y(i)$ is the average of the sum of the distance $d_j(i)$. As shown in Equation (8), the largest Lyapunov exponent λ is the slope of the straight line function of the variable i.

The evolution step i is set to 30, 60, 120 and 360 respectively. As shown in Figure 4, part of the largest Lyapunov exponent of the passenger flow series is greater than zero, indicating that the passenger flow series have chaotic characteristics.

V. Hybrid Model for Accurate Prediction of Passenger Flows

The quick growth of light rail networks and urban population explosion creates great pressure to the scheduling of subway operations. To make effective operation plans and alleviate the huge congestion problem in a light rail system with the regulation of passenger flows, planners need to know the time variations of short-term passenger flow between stations and between areas. A key problem to solve is to ensure the accurate prediction of the volume of the passenger flow between a pair of OD stations, based on which we can aggregate OD pairs to predict the total exit counts for a particular station or the total passenger flow between urban areas. In this section, we first introduce our problem system framework, and then detail our design of the prediction model.

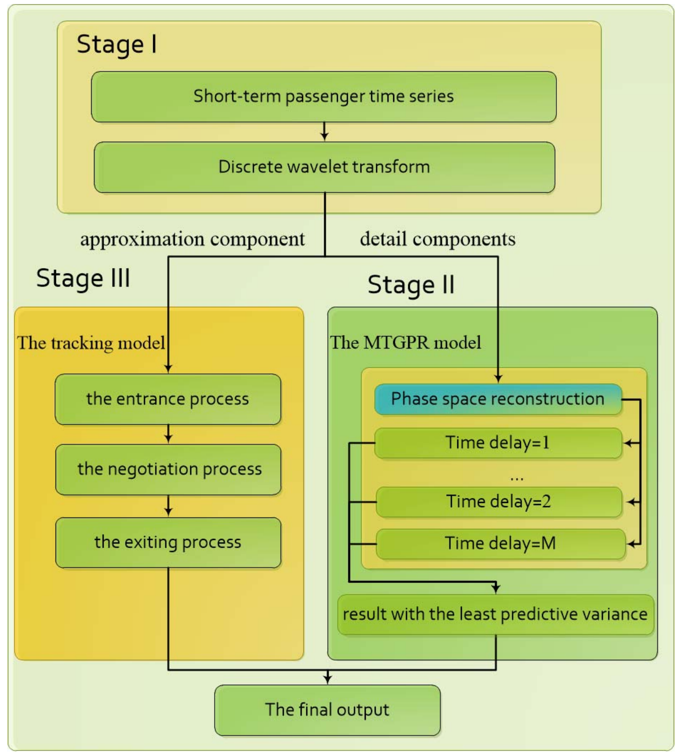


Fig. 5. The overall framework of the hybrid DTMGP model.

A. Problem and System Framework

We would like to predict the traffic volume between an OD pair corresponding to passengers entering from the subway station i and leaving from the station j. Let L_{it} be the number of passengers entering station S_i at time t (i.e., entering tap-in counts), N_{ijt} be the number of passengers entering from S_i and exiting from S_j at time t (i.e., tapping-out count). Given two sequences of observed passenger flow data $\{L_{it}, t = 1, 2, \cdots, T\}$ and $\{N_{ijt}, t = 1, 2, \cdots, T\}$, the prediction problem is to predict the value of $N_{ij(t+\Delta)}$, where Δ denotes the prediction horizon. According to the duration of Δ, we can divide the forecasting task into short, medium and long term. In this paper, we only discuss the short-term forecasting tasks, and set up the duration for each step of prediction to 1 min. The prediction for multiple steps ahead can be applied to predict passenger flows in the future, where we set up the ahead of time steps to 15, 20, 25, and 30 minutes respectively.

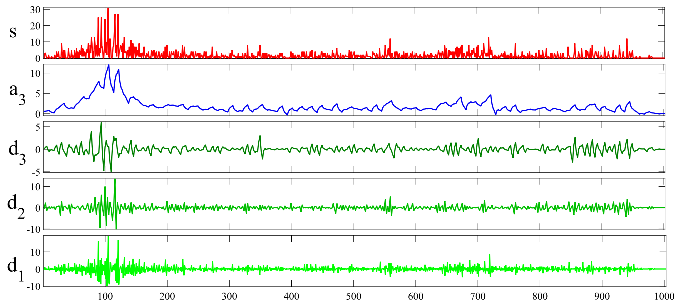


Fig. 6. Discrete wavelet decomposition

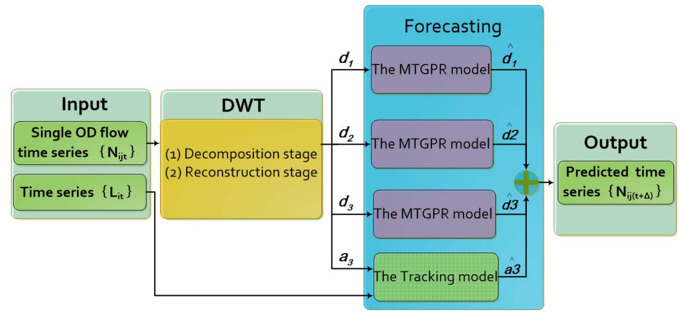


Fig. 7. Data processing procedure in the hybrid DTMGP model.

For more accurate passenger flow forecasting, we propose a hybrid model, named DTMGP. As shown in Figure 5, our prediction approach mainly contains three stages: (i) data preprocessing and decomposition, (ii) approximation component forecasting, and (iii) detail component forecasting.

In our model, we first apply discrete wavelet transform (DWT) to decompose raw traffic flow data into an approximation component and multiple detailed components. Then we apply a more efficient tracking model to forecast the approximation component. As the system is shown to have the chaotic and dynamics, rather than using the original time series data, we exploit the phase space method. Specifically, we reconstruct phase space from the detailed components and further propose a mixed time delay Gaussian process (MTGPR) model with a recursive strategy to forecast the detailed components. As the last step, we generate the final result with these forecasted components.

B. Data Preprocessing and Decomposition

A general traffic flow is mixed with smooth and continuous component and sudden changes due to rush hours or events. To better capture the dynamics of traffic flow, we exploit discrete wavelet transform to translate the data into frequency domain components, so that we can more accurately track the temporal evolution of the traffic flow at different time resolutions. The time resolution increases with the increase of the frequency level. Given a proper mother wavelet $\psi(t)$ and the decomposition level l, the discrete wavelet transform of a series $f(t)$ can be expressed as [40]

$$W(l,k) = \int_{-\propto}^{+\propto} f(t)\varphi^*_{l,k}(t)d_t, \varphi_{l,k}(t) = 2^{-j/2}\varphi(2^{-l}t - k) \tag{9}$$

where k is a time translation factor, $\varphi^*_{l,k}(t)$ is the complex conjugate of $\varphi_{l,k}(t)$, and $W(l,k)$ is the discrete wavelet coefficient at the level l and time k.

Now the original time series is represented as:

$$f(t) = \overline{C} + \sum_{l=1}^{L} W_l(t) \tag{10}$$

where the first term, $\overline{C}$, is the approximation component that captures the low frequency variation of the time series. The second term contains detailed components at levels $l = 1, 2, \cdots, L$. A higher level represents lower frequency and captures the remaining details left from the previous levels.

The input data of our hybrid model are two sequences of observed passenger flow data $\{L_{it}, t = 1, 2, \cdots, T\}$ and $\{N_{ijt}, t = 1, 2, \cdots, T\}$. We choose the Daubechies (Db) mother wavelet function and choose three levels of detailed decompositions. We first apply DWT to decompose the sequence $\{N_{ijt}, t = 1, 2, \cdots, T\}$ into one appropriation component (a_1) and three detailed components ($d1$, $d2$, $d3$), as shown in Figure 6. The red curve represents the sequence $\{N_{ijt}, t = 1, 2, \cdots, T\}$ within one day. The approximation component a_1 tracks the slow-varying long trend of the traffic volume. The curves d_1, d_2 and d_3 correspond to the high frequency data that fluctuate more dramatically in short-time scales, with d_1 representing the shortest time scale change. Then we take each of the three detailed components as the input to a MTGPR model for detailed component prediction. The sequence $\{L_{it}, t = 1, 2, \cdots, T\}$ and the appropriation component a_1 will be taken as the input to the tracking model for the appropriation component prediction. Finally, we sum up all these predicted components to the final result. Figure 7 illustrates the detailed data processing procedure in our model.

C. The More Efficient Tracking Model and Approximation Component Forecasting

We divide the traffic flow in a light rail system into three processes: the entrance process, the negotiation process and the exiting process. The entrance process tracks how many passengers enter a station to start a journey at a specific time. The negotiation process captures the number of passengers that enter from a specific origin and have stayed inside the system for a time duration. The exiting process captures the number of passengers that from a specific origin and exit from a destination station to end their journey, given the number of passengers inside the system and the length of their stay.

The Model for the Entrance Process: Silva *et al.* [34] model the expected value of L_{it} using $\theta_{L_{it}}$ and the history values $L_{i(t-w)}$, where $\theta_{L_{it}}$ represents an unconditional time-dependent mean that captures most of the variation of the data. However, our simulation results prove that this method may lead to large error. To better capture the dynamics of flow volume, we instead model the expected values of L_{it} under

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

the condition $t \geq 1$ and the set of given past entries as

$$E(L_{it}|PAST, L_{it} > 0) = (\theta_{L_{it}} + \sum_{w=1}^{W} \beta_{L_{wi}}(L_{i(t-w)} - \theta_{L_{i(t-w)}}))_+, \quad (11)$$

where $(x)_+$ means max(x,0), and

$$P(L_{it} > 0|PAST) \equiv \pi_{L_{it}} \quad (12)$$

In the above representation of the expected value of L_{it}, we modify $\theta_{L_{it}}$ by adding in the accumulative fluctuations of the history tap-in counts, with the difference at the time $t-w$ represented as $(L_{i(t-w)} - \theta_{L_{i(t-w)}})$. For simplicity, we define $L_{it} \equiv 0$ for $t < 1$.

The Model for the Negotiation Process: For each station S_i and time t, we provide a concise *presence table* to represent the number of passengers who enter the system via S_i and have not left the system by time $t-1$. The presence table captures the empirical distribution of passengers through the amount of time they have stayed inside the system. This empirical distribution is given in seven coarse time periods: [1, 10] min, [11, 20] min,. . ., [50, 60] min, and more than 60 min. For each station and time t, we have the vector $M_{it} \equiv (M_{it}^1, \ldots, M_{it}^7)$ represent the counts at these seven levels.

We model the temporal evolution of the entries of M_{it} through a cascade of nonparametric binomial regression models. We find that, for a given time t, the variation on $M_{it}^k(1 \leq k \leq 7)$ depends on $M_{i,t-1}^k + H_{it}^k - H_{it}^{k+1}$, where H_{it}^k represents the number of passengers who enter the system via the station S_i at time $t-1-(k-1)*10$ and have not left the system by time $t-1$. For example, for a time t of interest, M_{it}^3 represents the number of passengers who have stayed inside the system in the time bracket [21, 30]. It is not difficult to understand that these passengers are part of the ones who have been in the system in the time bracket [20, 29] at $t-1$, which is represented by $M_{i,t-1}^3 + H_{it}^3 - H_{it}^4$.

So M_{it}^k in a given day can be modeled as

$$M_{it}^k|PAST \sim Binomial(M_{i,t-1}^k + H_{it}^k - H_{it}^{k+1}, p_{it}^k), \quad H_{it}^k = L_{i,t-1-(k-1)*10} * q_{it}^k \quad (13)$$

where we define $M_{i,t-1}^0 \equiv L_{i,t-1}$. The parameter p_{it}^k represents the probability of entering the system via the station S_i during the period $[t-1-(k-1)*10, t-k*10]$ and staying in the system at time t. It is different for each station S_i and the time t of the day. The parameter q_{it}^k represents the probability of entering the system via the station S_i at $t-1-(k-1)*10$ and staying in the system for more than $(k-1)*10$ minutes, and it is fitted by the cubic spline smoothing.

The Model for the Exiting Process: According to history observations, we can calculate the median of the traveling time from the station S_i to the station S_j. The value of N_{ijt} will be impacted by the passenger numbers in the previous, current and next time brackets. For instance, given that the median of the traveling time from S_i to S_j is 35 minutes, the corresponding brackets are [21, 30], [31, 40], [41, 50]. The model for N_{ijt} is then

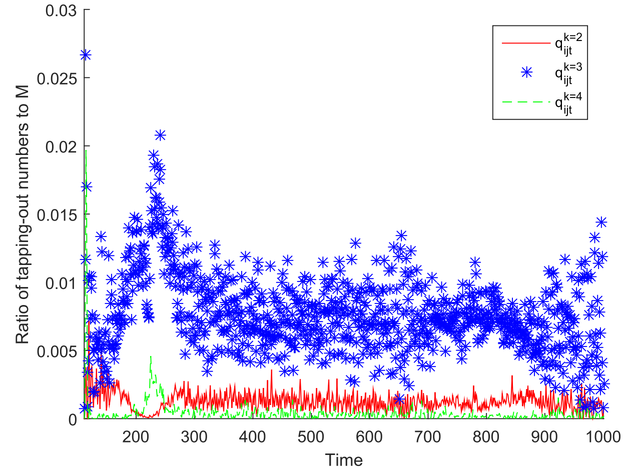


Fig. 8. The value of $q_{ijt}^{k=2}, q_{ijt}^{k=3}, q_{ijt}^{k=4}$ at different time t.

$$N_{ijt}|PAST = M_{it}^{k=2} \times q_{ijt}^{k=2} + M_{it}^{k=3} \times q_{ijt}^{k=3} + M_{it}^{k=4} \times q_{ijt}^{k=4} \quad (14)$$

where $q_{ijt}^{k=2}$ is the ratio of passengers from $M_{it}^{k=2}$ that leave from the station j at time t and contribute to N_{ijt}. In the Equation (14), we assign a different q_{ijt} to each M_{it}^k. To identify the traffic patterns, we conduct many experiments. Parameters q_{ijt}^k for a station pair $i-j$ are fitted using cubic spline smoothing. However, our experiment results demonstrate that the value of q_{ijt}^k varies largely for the same values of i, j, t. Figure 8 provides a visualization of q_{ijt}^k values, with $k = 2, 3, 4$. It is obvious that $q_{ijt}^{k=3}$ is often far greater than the other two values at a specific time t.

D. The MTGPR Model and the Forecast of Detailed DWT Components

As identified in Section IV, traffic flow data have strong chaotic characteristics. When handling this type of signals, traditional forecasting methods are often hard to meet the accuracy requirements. In addition, we usually need to reconstruct the state space first when making the prediction for the time series generated by a complex dynamic system. However, the passenger flow pattern in the complex dynamic system won't stay unchanged, and we cannot ensure that time delay selected for reconstructing a state space is optimal at any time.

Considering these factors, we propose an MTGPR model for more accurate traffic flow prediction. The schematic of multi-step-ahead (MS) prediction with an MTGPR model is shown in Figure 9, where the area between blue and green lines represent the forecasted values based on the Gaussian Process (GP) model. The larger the distance between the two lines, the greater the predicted variance. Given the embedding dimension d, we first select the total number of different time delays we will use in the prediction, M. We then reconstruct d-dimensional state space M times with different delay values, varying from 1 to M. For each space reconstructed, we make the prediction based on the GP model. Therefore, for the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

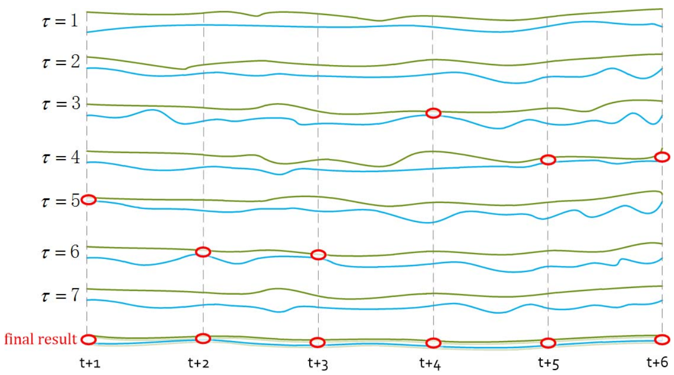


Fig. 9. Schematic of multi-step-ahead(MS) prediction with a mtgpr model.

prediction of y_{t+1}, we can obtain M pairs of predicted $mean_\tau$ and predicted $variance_\tau$. We put them in the set $\{(mean_\tau, variance_\tau)|\tau \in [1, M]\}$. We then select the pair with the least predicted variance as the final predicted result. For the prediction based on a phase space with a specific time delay, a red circle marks the smallest predicted variance. For a multi-step-ahead system, the uncertainty will increase as the number of steps looking ahead becomes larger. As one output of the classical GP model, the predicted variance reflects the uncertainty around the prediction mean. In order to minimize the uncertainty and improve the forecasting accuracy, we would like our prediction model to achieve the smallest predicted variance in each step.

Given a detailed component $\{y_i, i = 1, 2, \cdots, t\}$ extracted by DWT, we can obtain M input sets $\{(x_{i,\tau}, y_i), i = 1, 2, \cdots, t\}$ by reconstructing state spaces with different time delays τ (from 1 to M). Defining $y_i \equiv 0$ for $i < 1$, $x_{i,\tau} = (y_{i-1}, y_{i-\tau-1}, \cdots, y_{i-(d-1)\tau-1})$, $X_\tau = [x_{1,\tau}, x_{2,\tau}, \cdots, x_{t,\tau}]^T$, $Y = [y_1, y_2, \cdots, y_t]^T$, the MTGPR model for y_{t+1} is then:

$$\begin{aligned}\overline{y_{t+1}} &= E(y_{t+1}|X_{\hat{\tau}}, Y, x_{t+1,\hat{\tau}}),\\ \hat{\tau} &= \arg\min_{\tau\in[1,M]} var(y_{t+1}|X_\tau, Y, x_{t+1,\tau}),\\ &\qquad Y \sim N(0, K(X_\tau, X_\tau) + \sigma^2_{noise} I).\end{aligned} \tag{15}$$

where $E()$ and $var()$ represent the predicted mean and variance of forecasted values respectively based on the GP model. $K(X_\tau, X_\tau)$ is the covariance matrix, whose element $K_{ij} = k(x_{i,\tau}, x_{j,\tau})$ is the covariance function, usually given by a squared exponential form.

To reduce the overall processing time, we can use multi-thread or parallel computing to realize MTGPR. Suppose there is a h-step-ahead forecasting task, the algorithm can be executed iteratively h times with the recursive strategy. Though there exist dependency among different steps of looking ahead, parallel computing is just executed within each step. Given the historical observations $\{y_i, i = 1, 2, \cdots, t\}$, the time delay set $\{1, 2, \cdots, M\}$ and the embedding dimension d, there are four phases in each iteration:

1) *Initialization:* We initialize $M + 1$ processors/threads, with M processors responsible for making the prediction in parallel and the remaining one responsible for summarizing the results.
2) *Forecasting:* Each of the M processors makes one-step-ahead prediction based on the Gaussian process model and outputs a pair of predicted mean and predicted variance.
3) *Summarization:* The remaining one processor summarizes all the M results. The pair with the smallest predicted variance will be selected as the final result.
4) *Updating:* We update the input vector with the new predicted mean $\overline{y_i}$, where $i \in \{t + 1, t + 2, \cdots, t + h\}$. Then the program will continue to the next iteration.

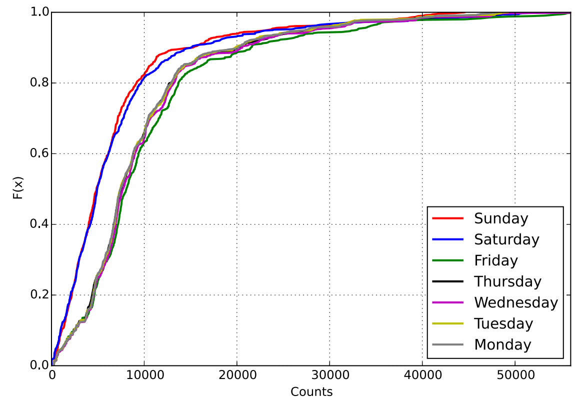


Fig. 10. Cumulative distribution function of exit counts aggregated per day.

VI. Experiments

We randomly select time series data from 100 pairs of OD stations to evaluate the performance of our prediction algorithm. As shown in Figure 10, the traffic patterns on weekdays and weekends are completely different. In order to reduce the impact of data difference on the experimental results, we divide the historical observations into two parts, one part from the weekdays and the other from weekends. The training process is independent in each part. We use data of the first 4 months as the training set and the later 1 month as the testing set.

To evaluate the performance of our hybrid model, we have performed two groups of extensive simulations. In the first group, we first evaluate the accuracy of our MTGPR model on making prediction for detailed components, and in the second group we evaluate the final forecasting performance of our hybrid model based on the original time series. We calculate the forecasting accuracy for each OD station pair and take the average to obtain the final forecasting accuracy.

A. Evaluation Metrics and Reference Models

We use four metrics to evaluate the forecasting performance of our model:

- DAY-MAE, the mean absolute error during the whole day.
- MR-MAE, the mean absolute error during morning rush hours.
- MPT-MAE, the mean absolute error of peak traffic forecasting during morning rush hours.
- DAYPT-MAE, the mean absolute error of peak traffic forecasting during the whole day.

MPT-MAE and DAYPT-MAE reflect the performance of our model on forecasting the congestion degree in a transportation

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

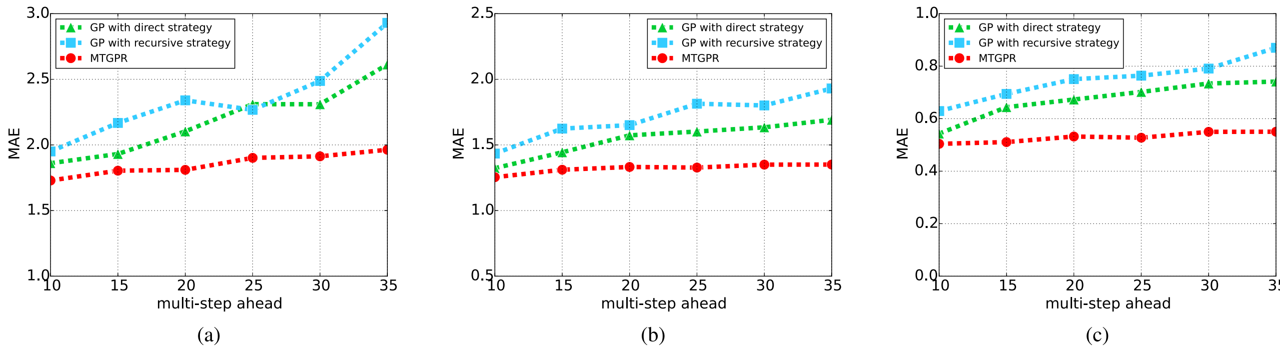


Fig. 11. (a) d1 component forecasting, (b) d2 component forecasting, (c) d3 component forecasting.

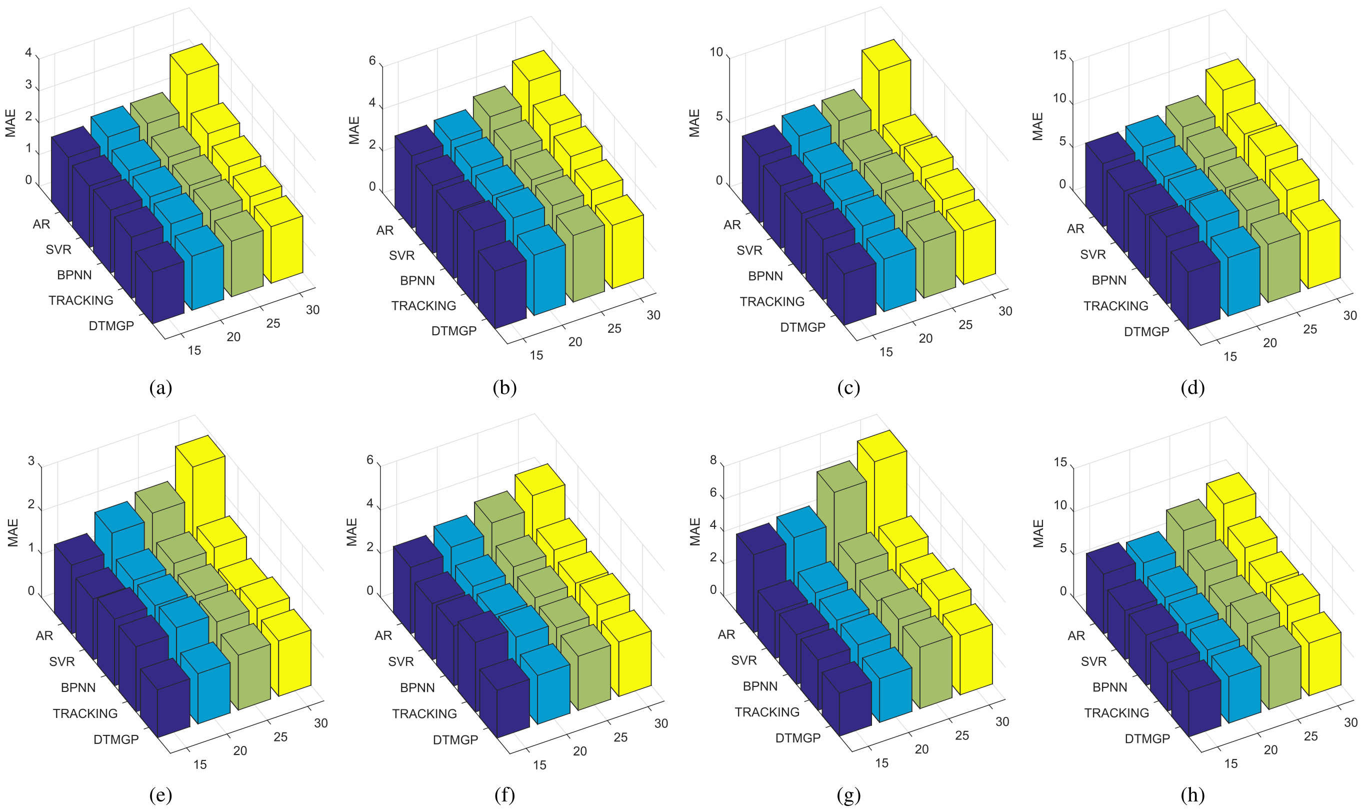


Fig. 12. Results predicted by different models: (a) DAY-MAE results comparison on weekdays, (b) DAYPT-MAE results comparison on weekdays, (c) MR-MAE result comparison on weekdays, (d) MPT-MAE results comparison on weekends, (e) DAY-MAE results comparison on weekends, (f) DAYPT-MAE results comparison on weekends, (g) MR-MAE result comparison on weekends, (h) MPT-MAE results comparison on weekends.

system. These metrics are defined as follows:

$$MAE = \frac{\sum_{t=t_{start}}^{t_{end}} |\hat{N_{ijt}} - N_{ijt}|}{t_{end} - t_{start} + 1} \tag{16}$$

where N_{ijt} represents the number of passengers that have entered the system from the station S_i and exits (tapping-out) from the station S_j at time t. $\hat{N_{ijt}}$ represents the forecasted value. For the metric DAY-MAE, $t_{start} = 1$ and $t_{end} = 1200$ (from 5:01 AM to 1:00 AM on the next calendar day). For the metric MR-MAE, $t_{start} = 141$ and $t_{end} = 240$ (from 7:21 AM to 9:00 AM). For the metric MPT-MAE, t include the moments at which the observed number of passengers soar to top30 during morning rush hours. For the metric DAYPT-MAE, t include the moments at which the observed number of passengers soar to top200 during the whole day.

We implement four reference forecasting models for the performance comparison.

1) *AR [41]:* AutoRegressive (AR) models are most widely studied because of their flexibility in modeling many stationary processes. In this study, we set the time lag I = 2, order p = 30.

2) *SVR [42]:* This approach reduces the over-fitting and computational costs inherent in traditional SVR, with the basic assumption that the most recent data samples provide more relevant information for forecasting.

3) *BPNN:* It is the most representative learning model for the ANN. BPNN in this study used 30 input neurons, 25 hidden nodes (three-layer) and 1 output neuron.

4) *Tracking [34]:* This model was designed to keep track of the number of passengers inside the transportation system.

B. Performance Results

1) Performance Comparison of Detailed Components Forecasting: In Figure 11(a-c), our MTGPR model is shown to be able to control the MAE on the d_1 component, d_2 component, and d_3 component at a very low level, while MAE under conventional Gaussian process model with direct strategy or recursive strategy are much higher. Even when raising the forecasting steps to 35, MAE values of our MTGPR model have very small increase, while the MAE of the other two models constantly increase with the prediction steps. This demonstrates the effectiveness of our model in minimizing the prediction error by constructing the phase space using the time delay with the minimum predicted variance in each step.

2) Performance Comparison of the Original Signals Forecasting: We carry out four series of simulations: 15-step ahead forecasting, 20-step ahead forecasting, 25-step ahead forecasting and 30-step ahead forecasting to evaluate the performance of our hybrid model. Figure 12 shows the forecasting performance of different models on weekdays and weekends. As shown in Figure 12, the forecasting using our hybrid model achieves the lowest error under all prediction steps studied.

With the number of prediction steps increased, the forecasting accuracy of the AR, SVR, BPNN and Tracking models deteriorate dramatically, while the forecasting accuracy of our hybrid model has very small increase. This indicates that our model does not have cumulative error and can better predict the traffic in advance, which will allow more time for efficient traffic scheduling. In addition, under the same steps, our hybrid model achieves on average $10\% - 30\%$ higher accuracy compared to other four models. It is worth noticing that, the DAYPT-MAE and MPT-MAE values by using the hybrid DTMGP model are only about $20\% - 45\%$ those of using other four models. This fact indicates that our model can better predict the overload degree in a transportation system, which will help alleviate the congestion. With the comprehensive consideration of spatial relationship among OD pairs, the temporal and frequency features of traffic, and the correlation among signals in the reconstructed phase space, our model can sense the small changes of external factors and timely adjust the predicted result.

Figure 12 (e-h) present the prediction accuracy by taking different methods on weekends. It can be easy to find that the forecasting accuracies of different methods are overall higher than those in Figure 12 (a-d). The reason for this phenomenon is that the passenger flow data on weekends are more stable compared with those on weekdays.

VII. Conclusion

In this paper, we propose a novel hybrid model to more accurately forecast the volume of passenger flows multi-step ahead. To the best of our knowledge, this is the first hybrid model to follow the evolution of traffic flow simultaneously from the temporal, frequency, OD spatial and historical probabilistic distribution perspectives. As one important part of the hybrid model, the MTGPR model proposed in this article can also effectively capture the process dynamics and make prediction with a smaller predictive variance compared with state of the art Gaussian process models. The simulation results demonstrate that our hybrid model can achieve on average $20\%-50\%$ higher accuracy compared to other models. In the future work, we also plan to explore methods to further improve forecasting accuracy by fully employing the graph structure of the road networks or light rail network.

References

[1] L. M. K. Boelter and M. C. Branch, "Urban planning, transportation, and systems analysis," *Proc. Nat. Acad. Sci. USA*, vol. 46, no. 6, pp. 824–831, 1960.

[2] J. R. Banavar, A. Maritan, and A. Rinaldo, "Size and form in efficient transportation networks," *Nature*, vol. 399, pp. 130–132, May 1999.

[3] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002.

[4] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[5] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, 1996.

[6] M. G. Karlaftis and E. I. Vlahogianni, "Memory properties and fractional integration in transportation time-series," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 4, pp. 444–453, 2009.

[7] B. Williams, "Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling," *Transp. Res. Rec. J. Transp. Res. Board*, no. 1776, pp. 194–200, 2001.

[8] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.

[9] I. Okutani, Y. J. Stephanedes, and F. Mannering, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, 1984.

[10] A. G. Hobeika and C. K. Kim, "Traffic-flow-prediction systems based on upstream traffic," in *Proc. Vehicle Navigat. Inf. Syst. Conf.*, Aug./Sep. 1994, pp. 345–350.

[11] M. S. Dougherty and M. R. Cobbett, "Short-term inter-urban traffic forecasts using neural networks," *Int. J. Forecasting*, vol. 13, no. 1, pp. 21–31, 1997.

[12] B. L. Smith and M. J. Demetsky, "Short-term traffic flow prediction: Neural network approach," *Transp. Res. Rec.*, no. 1453, pp. 98–104, 1994.

[13] M. Jun and M. Ying, "Research of traffic flow forecasting based on neural network," in *Proc. Int. Symp. Intell. Inf. Technol. Appl.*, Dec. 2008, pp. 104–108.

[14] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.

[15] H. Su, L. Zhang, and S. Yu, "Short-term traffic flow prediction based on incremental support vector regression," in *Proc. 3rd Int. Conf. Natural Comput.*, Aug. 2007, pp. 640–645.

[16] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6164–6173, 2009.

[17] R. K. Oswald, W. T. Scherer, and B. L. Smith, "Traffic flow forecasting using approximate nearest neighbor nonparametric regression," in *Final Project of ITS Center Project: Traffic Forecasting: Non-Parametric Regressions*, Dec. 2000.

[18] G. Yu, J. Hu, C. Zhang, and L. Zhuang, "Short-term traffic flow forecasting based on Markov chain model," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2003, pp. 208–212.

[19] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 1–9, Feb. 2014.

[20] Y. Xie, K. Zhao, Y. Sun, and D. Chen, "Gaussian processes for short-term traffic volume forecasting, " *Transp. Res. Rec., J. Transp. Res. Board*, no. 2165, pp. 69–78, Oct. 2010.

[21] C. Cheng, "Time series forecasting for nonlinear and non-stationary processes: A review and comparative study," *IIE Trans.*, vol. 47, no. 10, pp. 1053–1071, 2015.

[22] S. T. S. Bukkapatnam and C. Cheng, "Forecasting the evolution of nonlinear and nonstationary systems using recurrence-based local Gaussian process models," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 82, no. 5, p. 056206, 2010.

[23] W. Ni, S. K. Tan, W. J. Ng, and S. D. Brown, "Moving-window GPR for nonlinear dynamic system modeling with dual updating and dual preprocessing," *Ind. Eng. Chem. Res.*, vol. 51, no. 18, pp. 6416–6428, 2012.

[24] Y. Liu, H. Xiao, Y. Pan, D. Huang, and Q. Wang, "Development of multiple-step soft-sensors using a Gaussian process model with application for fault prognosis," *Chemometrics Intell. Lab. Syst.*, vol. 157, pp. 85–95, Oct. 2016.

[25] J. Hu and J. Wang, "Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression," *Energy*, vol. 93, pp. 1456–1466, Dec. 2015.

[26] S. Agarwal, P. Kachroo, and E. Regentova, "A hybrid model using logistic regression and wavelet transformation to detect traffic incidents," *IATSS Res.*, vol. 40, no. 1, pp. 56–63, 2016.

[27] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transp. Res. C, Emerg. Technol.*, vol. 21, no. 1, pp. 148–162, 2012.

[28] S. Wang, L. Yu, L. Tang, and S. Wang, "A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China," *Energy*, vol. 36, no. 11, pp. 6542–6554, 2011.

[29] X. Xing, X. Zhou, H. Hong, W. Huang, K. Bian, and K. Xie, "Traffic flow decomposition and prediction based on robust principal component analysis," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 2219–2224.

[30] Y. Zhang, Y. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 65–78, Jun. 2013.

[31] Y. Zhang, A. Haghani, and X. Zeng, "Component GARCH models to account for seasonal patterns and uncertainties in travel-time prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 719–729, Feb. 2015.

[32] J. Wang and Q. Shi, "Short-term traffic speed forecasting hybrid model based on Chaos–wavelet analysis-support vector machine theory," *Transp. Res. C, Emerg. Technol.*, vol. 27, pp. 219–232, Feb. 2013.

[33] F. Moretti, S. Pizzuti, S. Panzieri, and M. Annunziato, "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling," *Neurocomputing*, vol. 167, pp. 3–7, Nov. 2015.

[34] R. Silva, S. M. Kang, and E. M. Airoldi, "Predicting traffic volumes and estimating the effects of shocks in massive transportation systems," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 18, pp. 5643–5648, 2015.

[35] H. Froehling, J. P. Crutchfield, D. Farmer, N. H. Packard, and R. Shaw, "On determining the dimension of chaotic flows," *Phys. D, Nonlinear Phenomena*, vol. 3, no. 3, pp. 605–617, 1981.

[36] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A, Gen. Phys.*, vol. 45, no. 6, pp. 3403–3411, Mar. 1992.

[37] A. M. Fraser, "Information and entropy in strange attractors," *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 245–262, Mar. 1989.

[38] P. Grassberger, "Generalized dimensions of strange attractors," *Phys. Lett. A*, vol. 97, no. 6, pp. 227–230, 1983.

[39] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest lyapunov exponents from small data sets," *Phys. D, Nonlinear Phenomena*, vol. 65, nos. 1–2, pp. 117–134, 1993.

[40] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[41] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York, NY, USA: Springer, 2003.

[42] R. Khemchandani, Jayadeva, and S. Chandra, "Regularized least squares fuzzy support vector regression for financial time series forecasting," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 132–138, 2009.

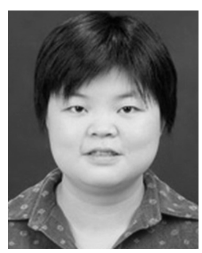
Zulong Diao received the B.E. degree in computer science and technology from Taiyuan University of Technology, Taiyuan, China, in 2013.

He is currently pursuing the M.E. degree in software engineering with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China. His research interests include deep learning, machine learning, complex networks, and human behavior dynamics.

Dafang Zhang received the Ph.D. degree in applied mathematics from Hunan University, China, in 1997. He is currently a Professor with the College of Computer Science and Electronic Engineering, Hunan University. His current research interests include dependable systems/networks, network security, network coding, traffic analysis, and network management.

Xin Wang received the B.S. degree in telecommunications engineering and the M.S. degree in wireless communications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 1990 and 1993, respectively, and the Ph.D. degree in electrical and computer engineering from Columbia University, New York, NY, USA.

She was a Member of Technical Staff in the area of mobile and wireless networking at Bell Labs Research, Lucent Technologies, NJ, USA, and an Assistant Professor with the Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY, USA. She is currently an Associate Professor with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY, USA. Her research interests include algorithm and protocol design in wireless networks and communications, mobile and distributed computing, and networked sensing and detection. He was a recipient of the NSF Career Award in 2005 and the ONR Challenge Award in 2010. She has served in executive committee and technical committee of numerous conferences and funding review panels, and serves as the Associate Editor of IEEE TRANSACTIONS ON MOBILE COMPUTING.

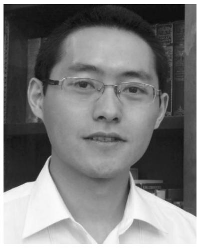


Kun Xie (M'05) received the Ph.D. degree in computer application from Hunan University, Changsha, China, in 2007. She was a Post-Doctoral Fellow with the Department of Computing, The Hong Kong Polytechnic University, from 2007 to 2010. She was a Visiting Researcher with the Department of Electrical and Computer Engineering, State University of New York, Stony Brook, from 2012 to 2013.

She is currently a Professor with Hunan University. Her research interests include wireless network and mobile computing, network management and control, cloud computing and mobile cloud, and big data. She has published over 70 papers in major journals and conference proceedings (including top journals IEEE/ACM TON, IEEE TMC, IEEE TWC, and IEEE TC, and top conferences INFOCOM, ICDCS, SECON, and IWQoS).

Shaoyao He received the B.S. degree from the College of Architecture, Hunan University, Changsha, China, in 1991.

She is currently a Professor with the College of Architecture, Hunan University. Her research interests include urban agglomeration, urban landscape planning, and building energy conservation technology.



Xin Lu received the B.S. degree in management science from Sichuan University, Sichuan, China, in 2006 and the Ph.D. degree in medical science from Karolinska Institutet, Stockholm, Sweden, in 2013.

He is currently an Associate Professor with the School of Business, Central South University, Changsha, China. His research interests include big data analytics, collective human behavior, complex networks, and emergency response.

Yanbiao Li was born in 1986. He received the Ph.D. degree in computer science from Hunan University, China, in 2016. In 2014, he was a Visiting Graduate Researcher with the Internet Research Lab, UCLA, for one year. He is currently a Post-Doctoral Fellow with Hunan University and a Visiting Research Associate with the Institute of Computing Technology, Chinese Academy of Science. His research interests include the future Internet architecture and high performance packet processing, especially on the design and implementation of fast and scalable data plane of NDN and SDN, respectively.